

# Microsatellite Data and Analysis

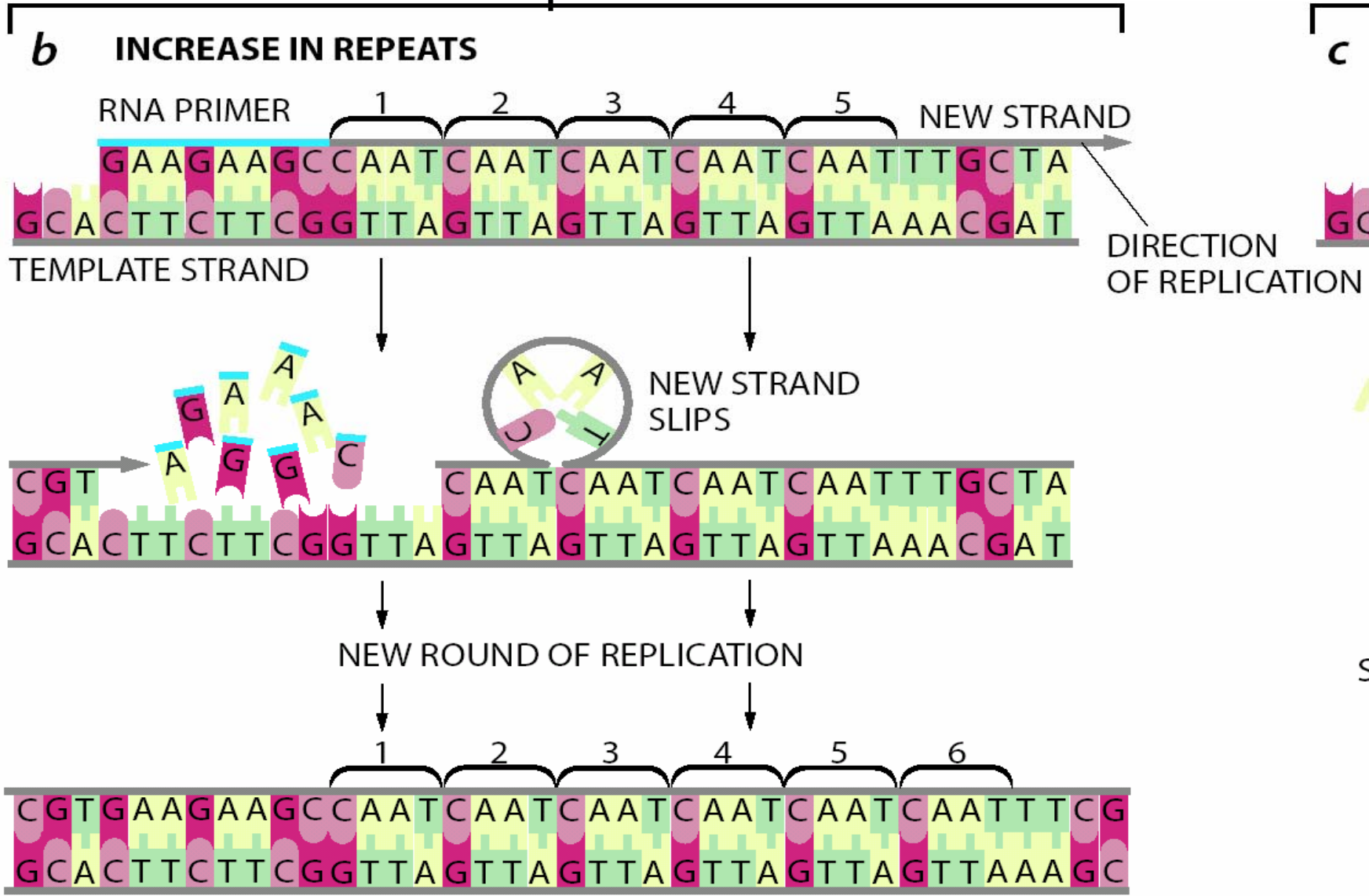
# Background

- Nuclear marker
- Simple sequence repeat (SSR)
- Simple tandem repeat (STR)
- Variation in number of tandem repeats
- Repeat usually 2-6 base pairs in length

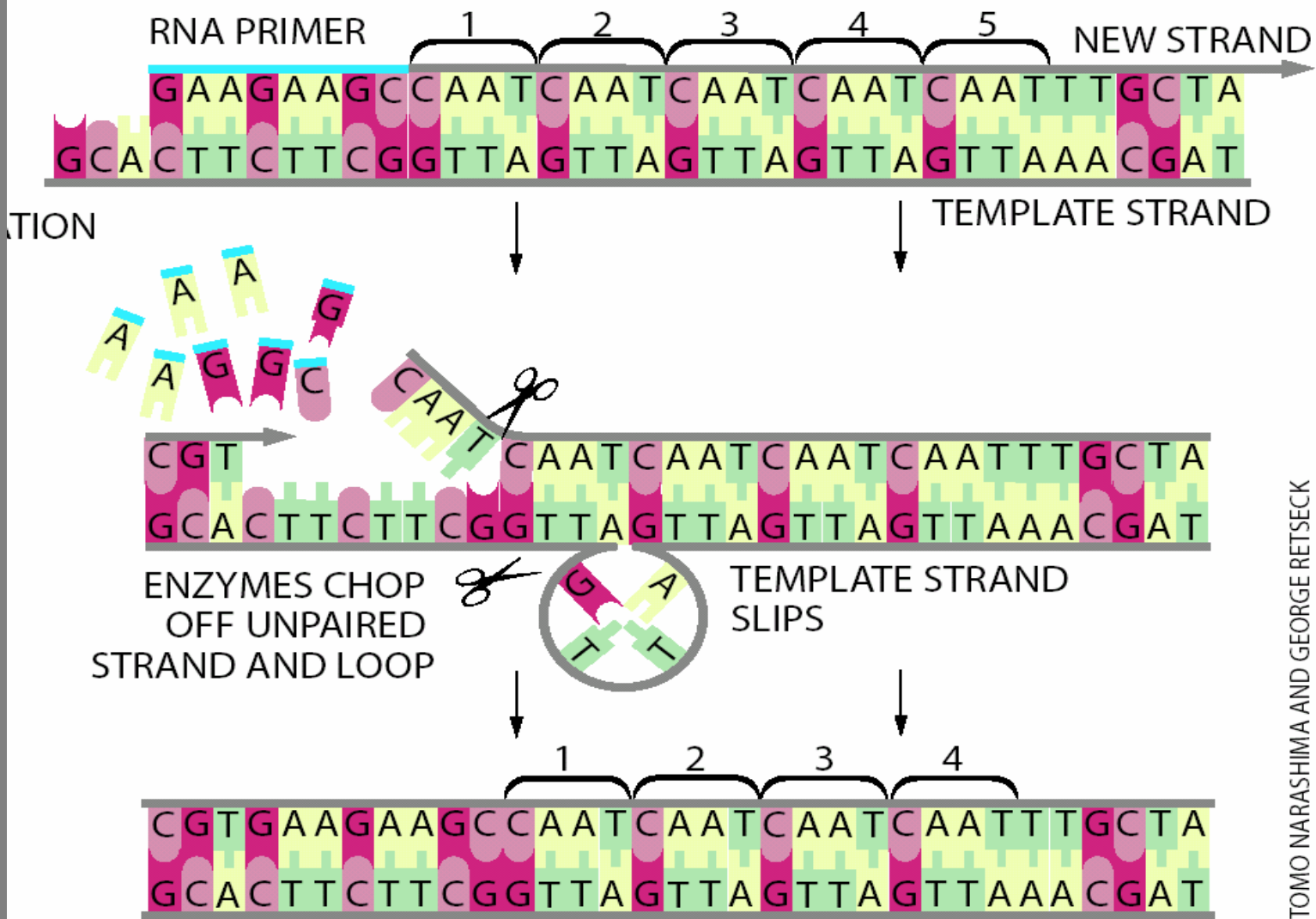
|              |     |     |     |     |     |              |
|--------------|-----|-----|-----|-----|-----|--------------|
| flanking DNA | CAT | CAT | CAT | CAT | CAT | flanking DNA |
| flanking DNA | GTA | GTA | GTA | GTA | GTA | flanking DNA |

# Mechanism of mutation

- DNA slippage
- DNA anneals out of register during replication
- Leads to expansion or contraction of microsatellite - most commonly a single repeat unit



**C DECREASE IN REPEATS**



TOMO NARASHIMA AND GEORGE RETSECK

# Why use microsatellites

- Abundant in eukaryotic genome
- Often relatively evenly spaced throughout genome
- Highly polymorphic
- Assumed to be selectively neutral
  - Tests for neutrality
- Commonly used in population comparisons

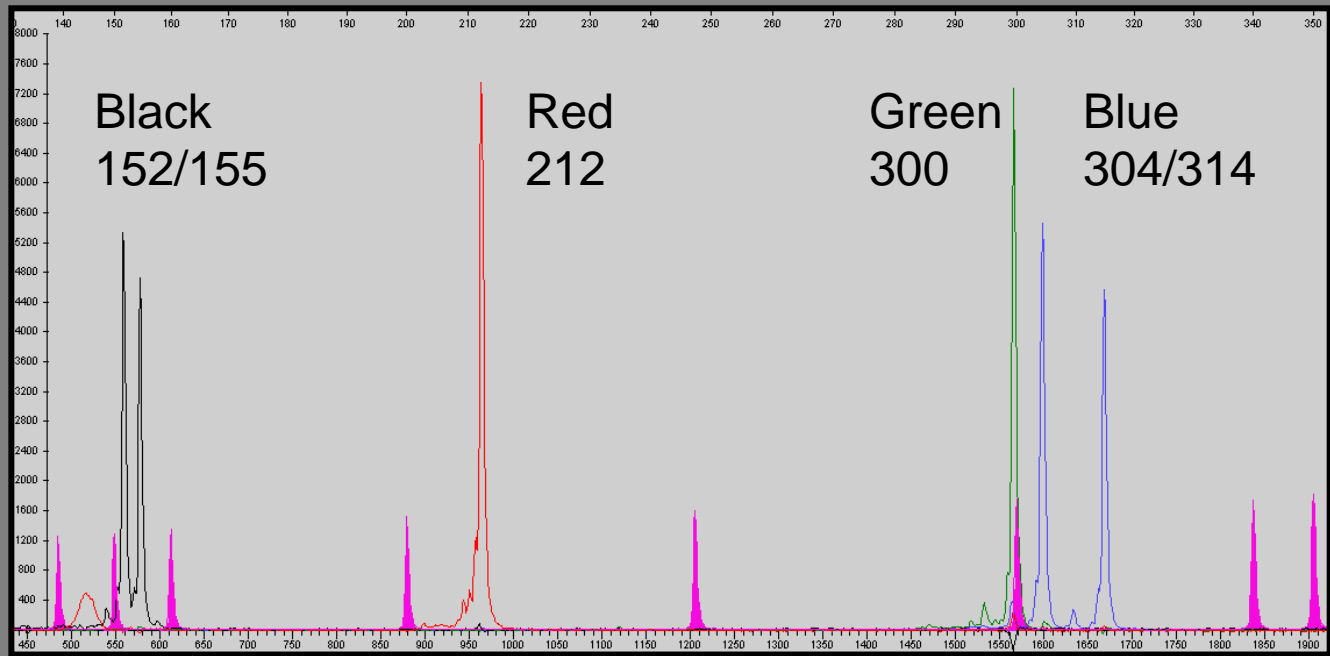
# Application: Individual & population level

- Individual identification
  - forensics, mark-recapture
- Genetic relatedness
  - paternity analysis
- Population genetics
  - breeding structure
- Conservation biology
  - between vs. within population genetic diversity

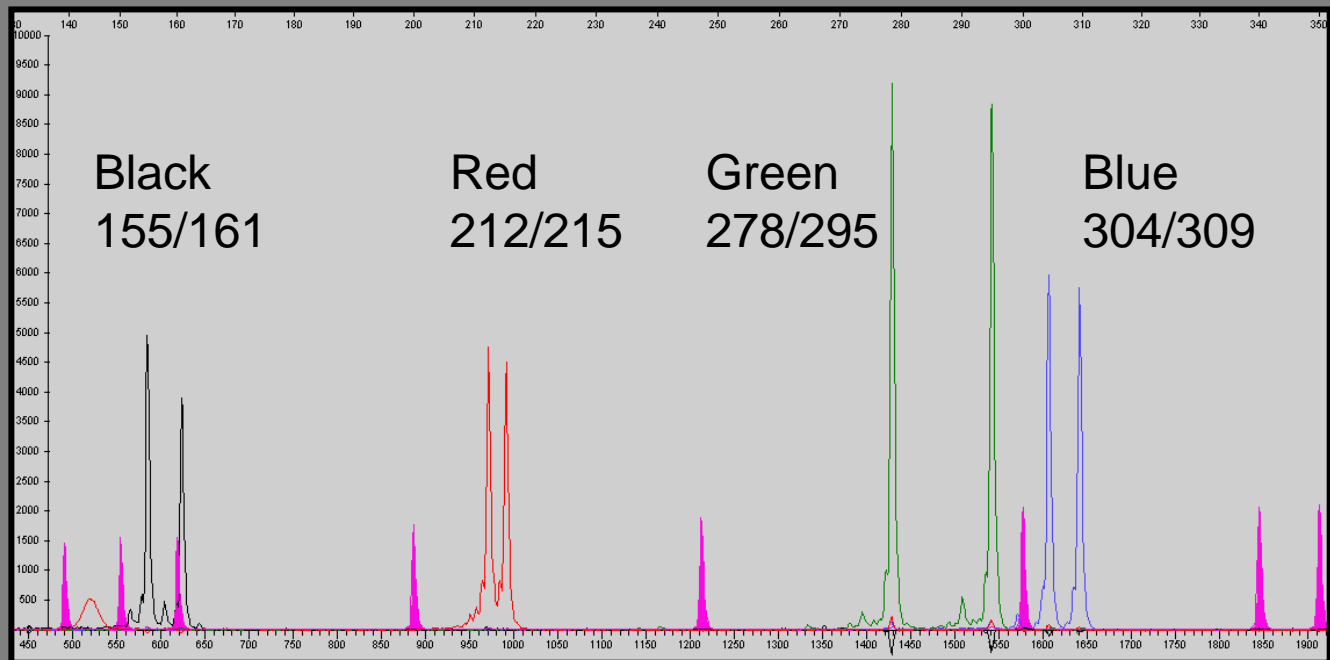
# Potential problems

- Need flanking sequence to develop primers
- Often significant effort in primer development for non-model organisms
- Null alleles
- Allelic drop-out
- Homoplasy
- Technical challenges
  - stutter
  - repeatability

Individual 5934



Individual 5910



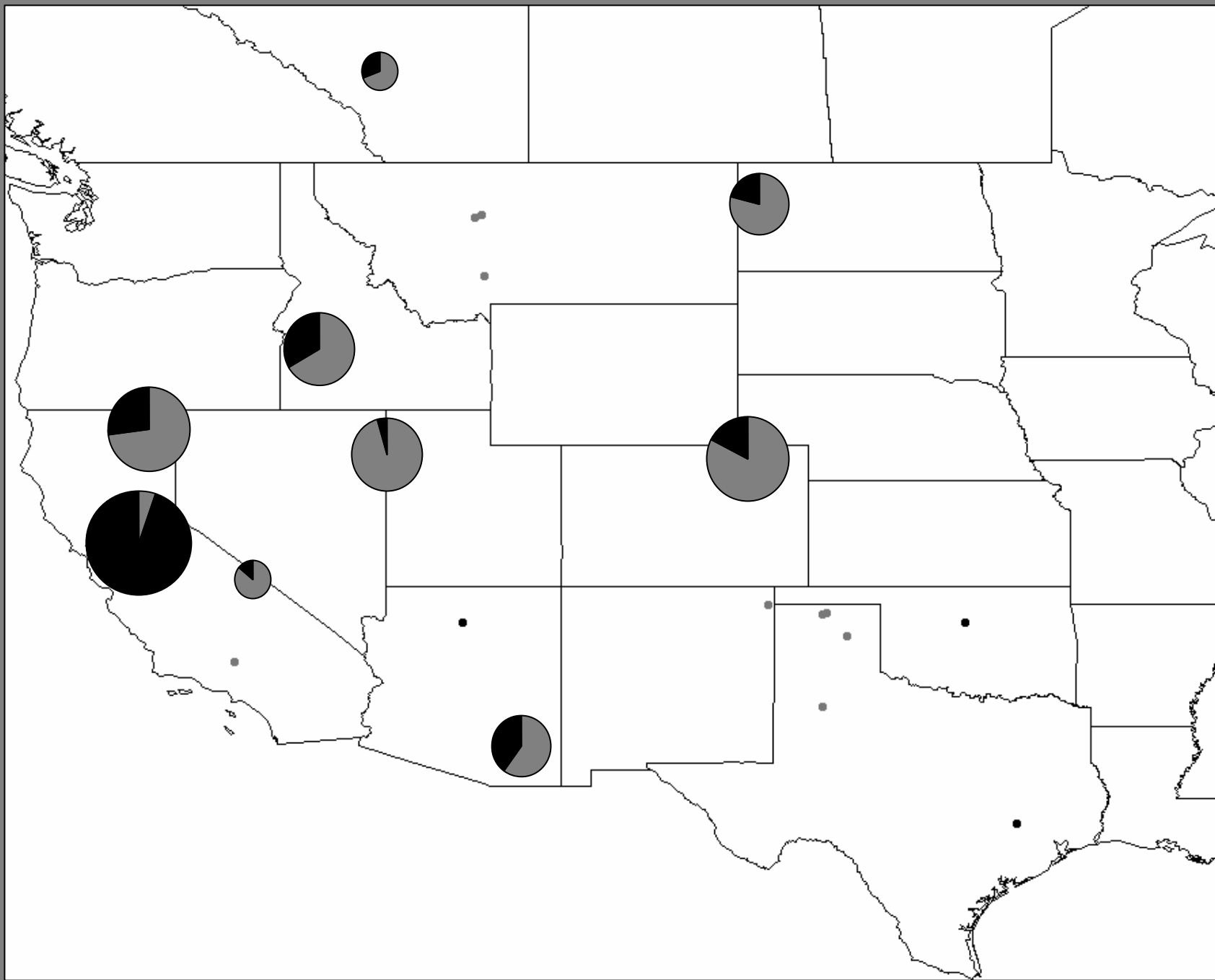
# Data files

| ID  | A110 |     | D122 |     | A204 |     | A317 |     | D210 |
|-----|------|-----|------|-----|------|-----|------|-----|------|
| 62  | 215  | 219 | 234  | 251 | 157  | 167 | 381  | 391 | 228  |
| 63  | 217  | 229 | 224  | 239 | 157  | 169 | 379  | 379 | 228  |
| 65  | 215  | 219 | 229  | 241 | 157  | 157 | 383  | 385 | 283  |
| 259 | 215  | 223 | 234  | 239 | 163  | 167 | 379  | 385 | 233  |
| 534 | 217  | 223 | 204  | 229 | 157  | 163 | 383  | 387 | 203  |
| 535 | 217  | 225 | 219  | 251 | 163  | 163 | 379  | 383 | 268  |
| 536 | 217  | 225 | 224  | 234 | 173  | 173 | 379  | 383 | 233  |
| 708 | 217  | 219 | 214  | 234 | 157  | 159 | 379  | 383 | 253  |
| 710 | 217  | 225 | 244  | 246 | 157  | 167 | 379  | 379 | 243  |
| 711 | 215  | 219 | 219  | 219 | 157  | 157 | 379  | 383 | 223  |
| 713 | 217  | 221 | 234  | 246 | 155  | 165 | 383  | 387 | 263  |
| 715 | 217  | 219 | 224  | 246 | 165  | 169 | 379  | 383 | 228  |
| 716 | 215  | 215 | 234  | 246 | 157  | 163 | 385  | 387 | 263  |
| 717 | 217  | 219 | 219  | 246 | 165  | 169 | 379  | 389 | 228  |
| 718 | 215  | 219 | 229  | 236 | 153  | 167 | 377  | 383 | 213  |
| 719 | 215  | 219 | 229  | 236 | 157  | 175 | 377  | 381 | 213  |
| 721 | 219  | 225 | 214  | 234 | 167  | 167 | 371  | 381 | 253  |
| 722 | 217  | 223 | 234  | 244 | 157  | 163 | 379  | 385 | 223  |
| 724 | 217  | 217 | 234  | 244 | 157  | 163 | 379  | 381 | 238  |

# Swainson's Hawk Example

- Do distinct populations exist within the range of Swainson's Hawks?



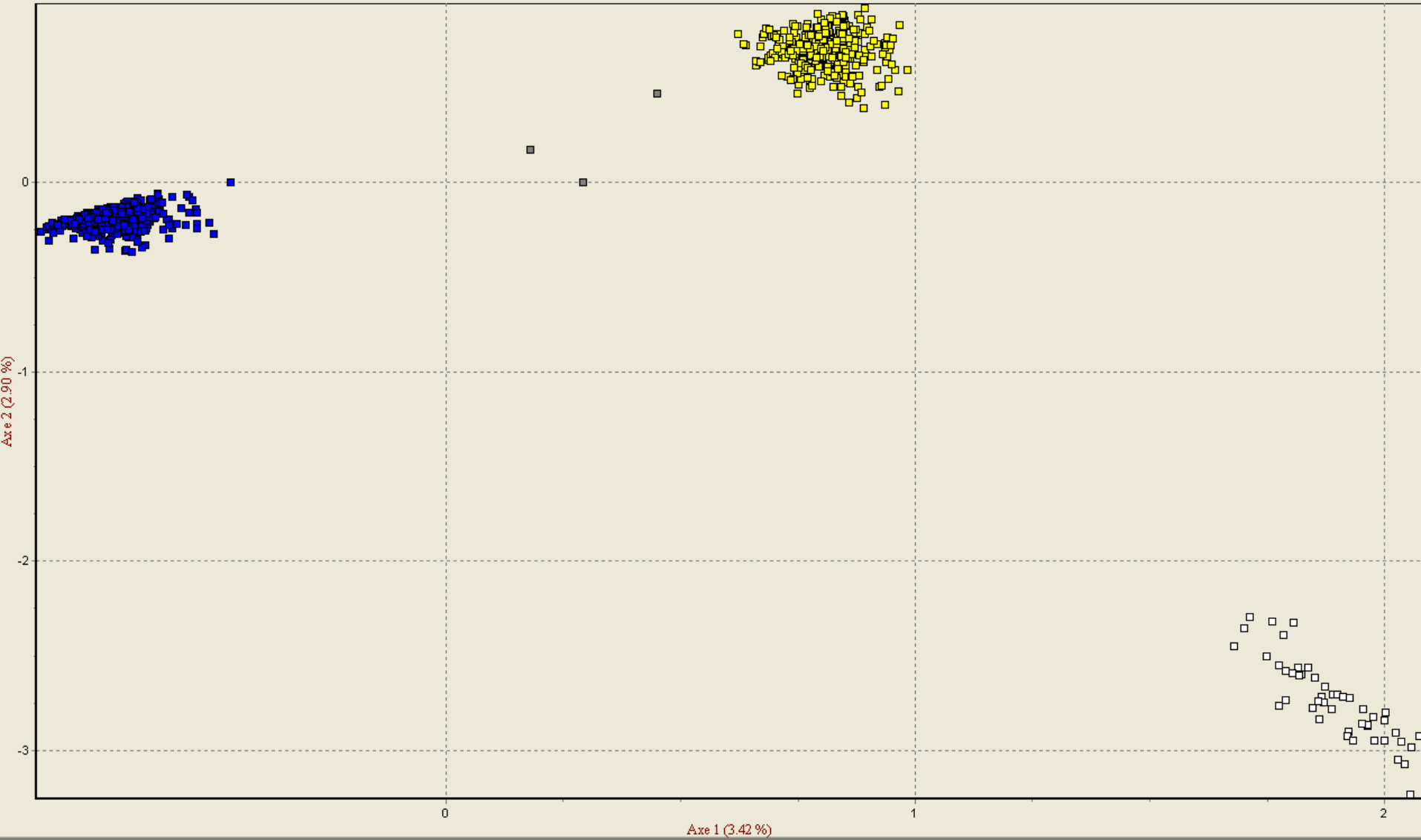


# Hybridization Among Raptors

- What is the origin of unusual raptor samples?



060801 SWHA RTHA WTHA no outlier genetix format.gtx



# Convert

The screenshot shows the main window of the CONVERT application. The title bar reads "CONVERT diploid genotypic data to other formats (GDA, GENEPOP, ARLEQUIN, POPGENE, MIC...". The menu bar includes "File". The main text area displays the following information:

```
*****
CONVERT.exe - version 1.31
- by Jeff Glaubitz - March 2005
- see CONVERT_help.doc for instructions and 'standard'
  input file format
- to begin, use "Load data file" (in File menu)
*****

CITATION:
Glaubitz JC (2004) CONVERT: a user friendly program to reformat diploid genotypic data for
commonly used population genetic software packages. Molecular Ecology Notes 4: 288-290.

Data file loaded:
\\vgl\lab\vgl-users\josh\Projects - Avian, Mammal,
diversity\CONVERT\061212 RTHA msat data 17 loci 3

- input data file in CONVERT format

Data Description:
17 loci RTHA data

Data file has 3 populations and 17 loci.
The total number of individuals is: 129
```

Overlaid on the main window is a dialog box titled "Choose desired format for file conversion". It contains a list of radio button options:

- GDA (nexus format: \*.nex)
- GENEPOP
- ARLEQUIN (\*.arp)
- POPGENE (\*.dat)
- MICROSAT (\*.mst)
- PHYLIP (allele frequency infile)
- STRUCTURE (\*.str)
- Produce table of allele frequencies

An "OK" button is located at the bottom of the dialog box.